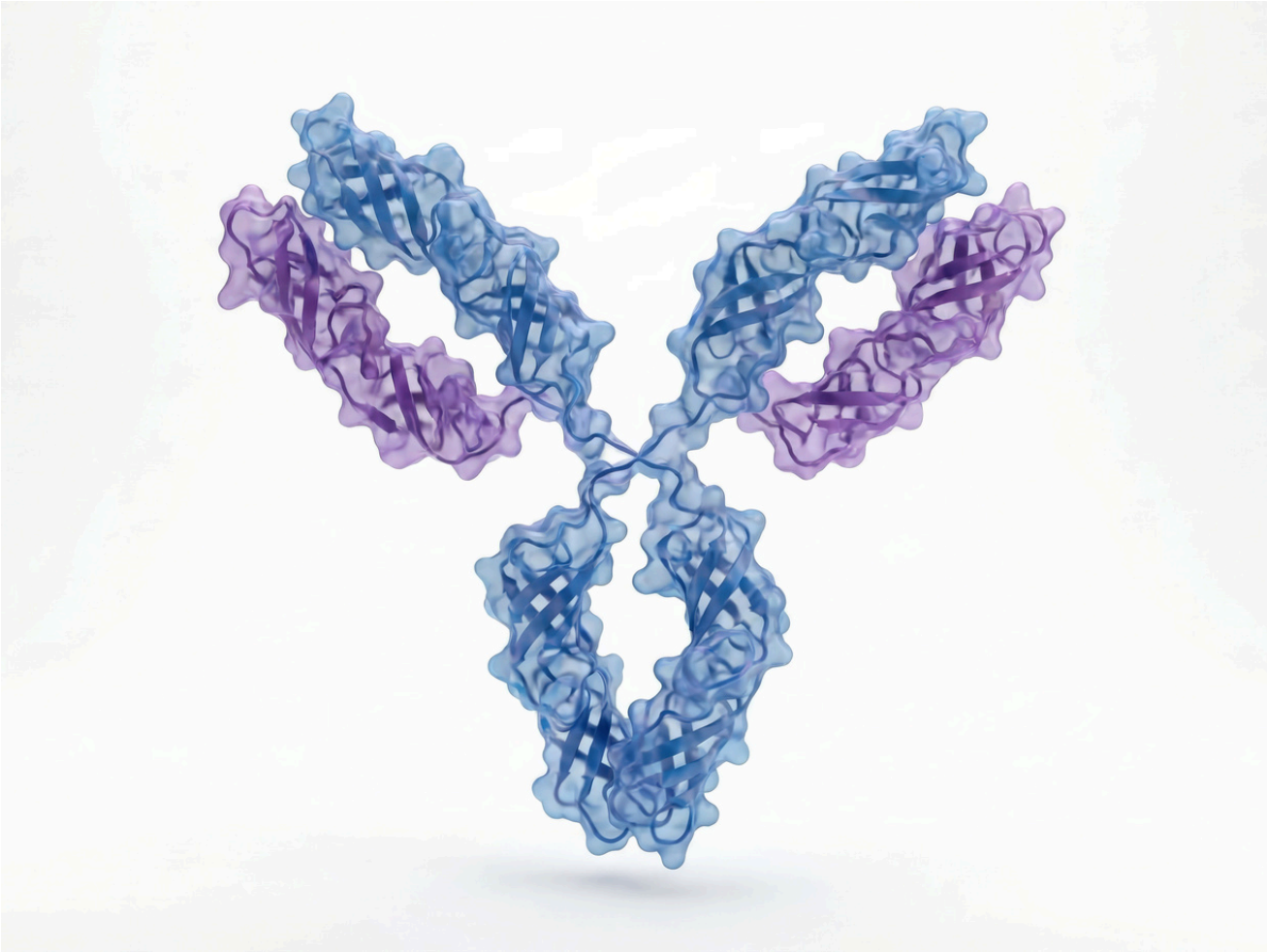
Ranomics

Engineering Robust Biologics: Developability, Function and AI

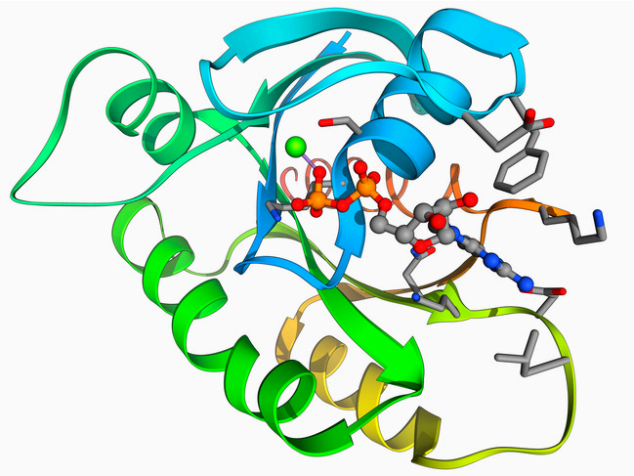
Biopharma and Biotechnological Applications



www.ranomics.com

The Paradigmatic Shift from Discovery to Developability

Historically, early biologics and enzyme discovery emphasized target binding and function, with developability often evaluated later in the funnel. This has led researchers to follow a classical framework of screening and characterizing high affinity binders. However, this "affinity-first" mindset can lead to risky candidates



The reality is that a high-affinity binder is a liability if it does not generate a biological response, survive manufacturing, and transport to patients. Teams are not simply longer for a needle in a haystack; they are looking for a needle that can be manufactured at production scale, remain stable at room temperature for weeks, and navigate the complex process of a high-concentration pre-filled syringe without aggregating.

In biologics development, many candidates that show excellent in vitro potency are ultimately discontinued because of poor developability. Common pitfalls include liabilities in solubility and aggregation, chemical stability, and non-specific or polyspecific binding, which can compromise formulation, manufacturability, pharmacokinetics, and safety even when target affinity is strong.

- Poor Solubility & Aggregation: High-concentration formulations (often required for subcutaneous delivery of mAbs) can lead to protein-protein interactions that cause precipitation or irreversible gelation.
- Chemical Instabilities: Post-translational modifications (PTMs) such as deamidation, oxidation, or isomerization can occur in the sequence, leading to a loss of potency over time.
- Polyspecificity: Many high-affinity binders can exhibit "sticky" behavior in vivo, binding non-specifically to unintended targets, which may impact clearance and safety when affinity is strong.

To de-risk these potential downstream hurdles, R&D teams must move toward front-loading developability, where biophysical fitness is screened simultaneously with functional activity.

By utilizing high-throughput platforms such as deep mutational scanning and cell display , a team can evaluate tens of thousands of variants across a "Developability Landscape." Instead of selecting a single lead and hoping it is stable, we identify the sequence-activity-stability relationships. This allows us to select candidates that reside in "fitness peaks" where high affinity overlaps with function, high stability and low poly-reactivity. These approaches, together with lower-cost DNA synthesis and AI-based developability prediction, enable the use of large datasets of both successful and failed variants to train models that highlight high-risk candidates before large-scale expression.

Selection Under Pressure—Next-Generation Display Platforms

To effectively de-risk a biologic, the selection system should closely mimic the molecule's future physiological and manufacturing environment. Traditional phage display, while powerful for its large library sizes, lacks the eukaryotic folding and post-translational machinery needed for more complex formats. For multispecifics, industrial enzymes, and CAR-T binders, shifting toward yeast and mammalian display enables earlier, more relevant developability screening.

Yeast Display: Quantitative Selection Workhorse

Saccharomyces cerevisiae is widely used to de-risk scFvs, VHHs, and industrial enzymes. Yeast cells are large enough for Fluorescence-Activated Cell Sorting (FACS), enabling quantitative selection based on multiple parameters. Dual labeling with an expression tag (e.g., c-Myc or HA) and a fluorescent ligand allows simultaneous measurement of display level (a proxy for folding and stability under defined conditions) and target binding. Yeast display also supports selection under stress. Libraries can be exposed to elevated temperatures, shifted pH, or organic solvents while still displayed on the cell surface, enriching variants that remain folded and functional under these conditions and flagging those more likely to fail in formulation or processing. To reduce polyspecificity, libraries can be incubated with "sinks" such as labeled serum or non-target paralogs and sorted to retain cells that bind the target while avoiding off-targets.

Mammalian Display: Capturing Human-Like Context

As biologics become more complex—such as bispecific T-cell engagers or glyco-engineered antibodies—mammalian display becomes increasingly important. Mammalian systems provide human-like post-translational modifications, including glycosylation patterns that influence Fc effector functions, FcRn binding, pharmacokinetics, and half-life. Screening libraries in this context reduces the risk that a candidate selected in a bacterial or yeast system will lose activity or change behavior once properly glycosylated.

For CAR-T development, mammalian display allows binders to be evaluated in the full receptor context, including the hinge, transmembrane, and signaling domains. This setup captures spatial constraints and signaling effects that simpler display systems can miss.

Deep Mutational Scanning as an Engine for AI

Modern display platforms can generate rich, high-dimensional data. By combining a cell-based functional assay with FACS (or other selection) and next-generation sequencing, deep mutational scanning (DMS) quantifies the effect of many single-point mutations in parallel. Comparing variant frequencies before and after selection yields an enrichment score for each mutation, which defines a fitness landscape highlighting residues that are essential for binding and those that tolerate change or create liabilities such as poor expression or aggregation risk. These experimentally derived fitness maps provide high-quality training data for machine learning models, including protein language models. Together, they support prediction of beneficial mutations and help de-risk sequence space beyond the specific variants physically tested in the library.

Beyond Binding—High-Throughput Functional De-risking

A frequent pitfall in biologics R&D is assuming that high binding affinity will automatically yield strong biological activity. In practice, the relationship between binding and function can be weak or non-linear, particularly for complex modalities such as multispecific antibodies, BiTEs, and CAR-Ts. To genuinely de-risk candidates, teams increasingly complement static binding measurements with custom high-throughput cell-based assays starting as early as lead optimization.

The functional gap: why binding is not enough

For a biologic to succeed, it must not only recognize its target but also drive the intended outcome. The outcome can range from triggering or blocking a signaling pathway, internalizing into the correct compartment, or catalyzing a reaction under relevant process conditions. In bispecific T-cell engagers, for example, potency depends on efficiently bridging T cells and target cells. Excessive tight binding to one partner can slow T-cell disengagement and reduce serial killing. Similarly, receptor-targeted molecules may show excellent affinity yet fail to induce the conformational change required for signalling.

Custom high-throughput reporter assays

Modern de-risking strategies use engineered reporter cell lines (such as Jurkat, CHO, or HEK293) that convert pathway activation into a measurable signal. By placing luciferase or fluorescent proteins under the control of pathway-specific response elements, researchers can screen large variant panels for their effects on a defined signaling axis in pooled or plate-based formats. Dual-reporter systems enable “coincidence detection,” where one readout tracks desired target activity (for example, pathway activation in green) and a second readout tracks stress or toxicity (for example, a red stress reporter), allowing selection of variants that are both potent and selective.
Functional de-risking for specialized modalities

For CAR-T and related cellular therapies, functional de-risking focuses on readouts such as target cell killing and cytokine production across libraries of CAR constructs. High-throughput flow cytometry and co-culture assays make it possible to evaluate many designs for their ability to eliminate tumor cells, including those with low antigen density. In enzyme engineering, function is defined by substrate turnover under application-relevant conditions; droplet microfluidics and colorimetric plate-based assays support screening thousands of variants per day for catalytic efficiency and robustness, including performance at low substrate concentrations or in the presence of detergents, chaotropes, or organic solvents.

The role of AI in functional prediction

The large functional datasets generated by these assays can be used to train models that approximate an “activity landscape” over sequence space. Structural prediction tools like AlphaFold help assess whether a sequence is likely to fold, but they do not reliably predict potency or pathway-level effects on their own. By feeding high-throughput assay results into machine learning pipelines, researchers can begin to prioritize variants in silico, focusing experimental effort on sequence changes most likely to enhance the desired biological response rather than merely improving target binding.

AI/ML Integration for Predictive Modeling

Modern AI tools have evolved beyond simple filtering of existing libraries to actively generating promising candidates.

De novo binder design: Tools like RFdiffusion generate novel protein binders for challenging epitopes and extracellular domains by optimizing residue placement to match target pocket geometry.

Protein language models (PLMs): Models such as ESM-2 and ProtT5, trained on billions of natural protein sequences, capture the "grammar" of folding and stability. These tools can identify sequence features associated with developability risks—like deamidation-prone motifs (NG/NS) or hydrophobic patches—for experimental follow-up, even before synthesis.

Computational Developability Platforms

Integrated computational workflows combine sequence, structure, and biophysical predictions to assess candidates earlier in development.

In silico formulation: Machine learning models trained on molecular dynamics simulations predict high-concentration behaviors such as viscosity and aggregation propensity, helping formulation teams evaluate subcutaneous delivery feasibility during lead selection.

Immunogenicity assessment: Tools like AbImmPred scan sequences for potential T-cell epitopes that bind MHC-II molecules, enabling targeted de-immunization by swapping problematic residues while preserving binding affinity.
Closing the Loop: Lab-in-the-Loop Workflows

The greatest value emerges from iterative cycles linking computation and experiment.

Active learning: AI proposes targeted variant sets (e.g., 1,000 designs), which are rapidly screened using yeast/mammalian display or functional assays. The resulting data—both high-performing hits and informative failures—feeds back to refine the model, accelerating convergence on optimal candidates.

Collaborative model training: Emerging federated learning approaches enable multiple organizations to improve shared developability models without exchanging proprietary sequence data, potentially benefiting the broader field.

The Roadmap—A Checklist for R&D Leaders

In 2026, the competitive edge belongs to R&D teams who can compress timelines without increasing risk. Use this checklist to audit your current biologics pipeline and identify gaps in your de-risking strategy.

1. Library Quality & Selection Strategy

- [ ] Beyond Theoretical Diversity: Is your library construction method biased toward functional sequences? (Ensure >90% of your library is in-frame and foldable).
- [ ] Fidelity of the Host System: Are you using mammalian display for complex multispecifics to ensure human-like post-translational modifications (PTMs) from day one?
- [ ] The Normalization Factor: Does your screening workflow (e.g., FACS) normalize for display/expression levels to ensure affinity results are not confounded by titer?

2. Deep Functional De-risking

- [ ] DMS Implementation: Are you utilizing Deep Mutational Scanning (DMS) to map the fitness landscape of your leads?
- [ ] Negative Selection: Have you integrated "sink" molecules or decoy receptors into your high-throughput (HTP) assays to eliminate off-target binders early?
- [ ] Stress-Test Integration: Are thermal, chemical, or pH-based challenges performed during selection, rather than months later in pre-formulation?

3. AI and the Digital Twin

- [ ] The Ground Truth Loop: Does your data flow directly into a machine learning model to refine the next round of library design? (Lab-in-the-Loop).
- [ ] Predictive Liability Mapping: Do you use Protein Language Models (PLMs) to scan for in silico liabilities (deamidation, glycosylation motifs, immunogenic epitopes) before wet-lab validation?
- [ ] Negative Data Utilization: Are your "failed" variants being sequenced and used to train your AI, or is that data being discarded?

4. Operational & Manufacturing Readiness

- [ ] Scale-up Proxy: Does your custom cell-based assay reflect the physiological disease state (e.g., primary cell-like signaling) rather than a simple over-expressed reporter line?
- [ ] High-Concentration Simulation: Are you using AI to predict viscosity and aggregation propensities for subcutaneous delivery early in lead optimization?
- [ ] Regulatory Preparedness: Are you documenting your AI-driven selection rationales to meet 2026 standards for "Explainable AI" in regulatory submissions?

Further Readings

Developability Assessment & Early-Stage Screening

1. Zhang D, et al. "Developability assessment at early-stage discovery to enable development of antibody-derived therapeutics." Antibody Therapeutics 6(1):13-25, 2023.
2. Cunningham O, et al. "Polyreactivity and polyspecificity in therapeutic antibody candidates." mAbs 13(1), 2021.
3. Herling TW, et al. "Nonspecificity fingerprints for clinical-stage antibodies in serum." PNAS 2023.
4. Yu X, et al. "Protein language models enable prediction of polyreactivity of monoclonal antibodies." Antibody Therapeutics 7(3):199, 2024.
5. Greaney AJ, et al. "Deep mutational learning for the selection of therapeutic antibodies." Nature Biomedical Engineering 2025.

Display Technologies (Yeast, Mammalian, DMS)

1. Boder ET, Wittrup KD. "Yeast surface display for screening combinatorial polypeptide libraries." Nature Biotechnology 15(5):553-7, 1997.
2. Chao G, et al. "Isolating and engineering human antibodies using yeast surface display." Nature Protocols 1(2):755-68, 2006.
3. Gai SA, Wittrup KD. "Yeast surface display for protein engineering and characterization." Current Opinion in Structural Biology 17(4):458-64, 2007.
4. Cherf GM, Cochran JR. "Applications of yeast surface display for protein engineering." Methods in Molecular Biology 1319:155-75, 2015.
5. Slavny P, et al. "Advancements in mammalian display technology for therapeutic antibody discovery and engineering." Current Opinion in Biotechnology 2024.
6. Wagner DL, et al. "A high-throughput screening strategy identifies CAR-T designs with enhanced potency." Nature Biotechnology 2022.

High-Throughput Functional Screening

1. Mazor Y, et al. "Impact of intrinsic affinity on functional binding and biological activity of EGFR-blocking antibodies." mAbs 6(4):1020-9, 2012.
2. Wang Y, et al. "High-throughput functional screening for next-generation therapeutic antibodies." Science Advances 7(24):eabe3839, 2021.
3. Ryoo H, et al. "Advances in high throughput cell culture technologies for stem cell differentiation, drug discovery & development." Advanced Healthcare Materials 13(3):2302485, 2023.

AI/ML Integration & Computational Design

1. Watson JL, et al. "De novo design of protein structure and function with RFdiffusion." Nature 620(7976):1089-1100, 2023.
2. Yang KK, et al. "Improved protein binder design using β-pairing targeted RFdiffusion." Nature 2026 (PubMed 41519838).
3. Morehead AK, et al. "Atomically accurate de novo design of antibodies with RFdiffusion." bioRxiv 2024 (PMC10983868).
4. Rives A, et al. "Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences." PNAS 118(15):e2016239118, 2021.
5. Hie BL, et al. "Efficient evolution of human antibodies from general protein language models." Nature Biotechnology 2024.